

Evolution of sociality in spiders leads to depleted genomic diversity at both population and species level

Settepani, V.^{*}, Schou, M. F.^{*}, Greve, M.[†], Grinsted, L.[‡], Bechsgaard, J.^{*}, Bilde, T.^{*}

^{*}Department of Bioscience, Aarhus University, Ny Munkegade 116, building 1540, 8000 Aarhus C, Denmark

[†] Department of Plant Science, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

[‡]School of Life Sciences, University of Sussex, John Maynard Smith Building, Brighton BN1 9QG, UK

Keywords: evolutionary dead end, genetic diversity, *Stegodyphus*, inbreeding, RAD sequencing, demographic modelling

Corresponding author: Trine Bilde, Ny Munkegade 116, building 1540, 8000 Aarhus C, Denmark, Tel.: +4560202702, Fax: +45 87150201, Email: trine.bilde@bios.au.dk

Short title: Genomic consequences of social evolution

Abstract

Across several animal taxa, the evolution of sociality involves a suite of characteristics, a ‘social syndrome’, that includes cooperative breeding, reproductive skew, primary female biased sex-ratio, and the transition from outcrossing to inbreeding mating system, factors that are expected to reduce effective population size (N_e). This social syndrome may be favoured by short-term benefits but come with long-term costs, because the reduction in N_e amplifies loss of genetic diversity by genetic drift, ultimately restricting the potential of populations to respond to environmental change. To investigate the consequences of this social life form on genetic diversity, we used a comparative RAD-sequencing approach to estimate genome-wide diversity in spider species that differ in level of sociality, reproductive skew, and mating system. We analysed multiple populations of three independent sister-species pairs of social inbreeding and subsocial outcrossing *Stegodyphus* spiders, and a subsocial outgroup. Heterozygosity and within population diversity were 6-10 fold lower in social compared to subsocial species, and demographic modelling revealed a tenfold reduction in N_e of social populations. Species-wide genetic diversity depends on population divergence and the viability of genetic lineages. Population genomic patterns were consistent with high lineage turnover, which homogenizes the genetic structure that builds up between inbreeding populations, ultimately depleting genetic diversity at the species level. Indeed, species-wide genetic diversity of social species was 5-8 times lower than that of subsocial species. The repeated evolution of species with this social syndrome is associated with severe loss of genome-wide diversity, likely to limit their evolutionary potential.

Introduction

Understanding processes that influence the genetic diversity of natural populations is central for predicting the potential of species to respond to environmental change, i.e. their evolutionary potential. In large populations, mutation, selection and migration influence genetic polymorphisms, but in reality population sizes are finite and affected by stochastic fluctuations in allele frequencies (Ellegren & Galtier 2016; Wright 1931). Change in allele frequencies across generations due to random sampling, or genetic drift, is determined by effective population size (N_e). Because of the inverse relationship between N_e and genetic drift, changes in N_e over time are expected to strongly affect population genetic diversity (Alcala & Vuilleumier 2014). In addition to census size, N_e is influenced by the ecology and demography of species, for example life history, dispersal, habitat fragmentation, and mating system, and by population dynamic processes such as extinction/colonisation events that may result in severe population fluctuations (Charlesworth 2009; Ellegren & Galtier 2016; Lourenço *et al.* 2017). The demographic history of species is influenced by biotic and abiotic environmental changes, and ecological drivers of demographic fluctuations may for instance result in population bottlenecks or founder effects that rapidly deplete genetic diversity due to drift. Reproductive mode and life history is highly correlated with genetic diversity in the way that species with few offspring, parental care, and long generation time, are less diverse (Romiguier *et al.* 2014a). These factors may act differentially on N_e , but if their cumulative effect reduces N_e , we expect reduced genetic diversity and increased structure within and among populations (Charlesworth 2003; Hamrick & Godt 1996; Leffler *et al.* 2012; Romiguier *et al.* 2014a).

The evolution of cooperative breeding is widespread across the animal kingdom, and this social lifestyle is associated with several traits that act to reduce N_e . Cooperative breeding frequently

involves reproductive division of labour where only one or a few females reproduce within family groups, for example in insects (Wilson 1971), birds (Cockburn 1998), mammals (Lukas & Clutton-Brock 2012), fish (Taborsky 2009), and arachnids (Lubin & Bilde 2007). In addition, some social species show female-biased sex ratio (Bourke 1997; Hamilton 1967; Nomura & Takahashi 2012; West 2009), and the social insects (bees, ants, and wasps, thrips and some bark beetles) are haplodiploid, which further decreases N_e (Hedrick & Parker 1997). Most social species maintain an outcrossing mating system through sex-biased dispersal (Clobert *et al.* 2001; Pusey 1987; Tabadkani *et al.* 2012; Wilson 1971), and thereby reduce the risk of inbreeding. In contrast, sociality in a number of taxa involves a regular inbreeding or even an obligatory inbreeding mating system. In these systems, social groups form by philopatry through the retention of offspring in the natal nest, which results in kin-group structuring within populations (Bilde *et al.* 2005). Benefits conveyed by group living or cooperation acting in concert with ecological constraints may further select for limited dispersal (Hatchwell & Komdeur 2000; Koenig *et al.* 1992; Thorne 1997), and increase the probability of reproducing with relatives (Bilde *et al.* 2005). This results in the transition from outcrossing to inbreeding mating system. Inbreeding may have severe detrimental effects on viability through the unmasking of recessive deleterious alleles (Charlesworth & Willis 2009), although purging may remove some of the genetic load (Crnokrak & Barrett 2002). In addition, the correlation between the parental alleles of an individual caused by inbreeding reduces N_e (Charlesworth 2009). The combined effects of inbreeding, female biased sex ratio, and reproductive skew are expected to result in severe reduction of N_e and random loss of alleles through genetic drift (Charlesworth 2003; Charlesworth & Willis 2009), thereby substantially depleting population genetic diversity (Charlesworth & Wright 2001; Wright *et al.* 2008). Social species with these characteristics that we refer to as a ‘social syndrome’ (Bilde & Lubin 2011),

include thrips (Chapman *et al.* 2000), beetles (Keller *et al.* 2011), aphids (Johnson *et al.* 2000), mole rats (Reeve *et al.* 1990), and spiders (Lubin & Bilde 2007).

Theoretically, inbreeding is expected reduce N_e up to 50 % (self-fertilization), while the effects of sex ratio bias and reproductive skew further reduces N_e , depending on the strength of the bias and skew respectively (Wright 1931). For example, if a population is female biased with 9:1, N_e is reduced to 36%, while N_e of a population with an equal sex ratio but where only half of one sex reproduces is reduced to 67%. Regular inbreeding is often associated with life history characteristics that promote more severe population size fluctuations (Ingvarsson 2002; Schoen & Brown 1991) due to frequent cycles of extinction-recolonization events (Charlesworth & Wright 2001), and increased extent of linkage disequilibrium and genetic hitchhiking (selective sweeps and background selection) (Nordborg 2000). In some species, new populations are formed by propagule dispersal of mated females, which leads to recurrent founder events further decreasing genetic diversity of newly established populations. The effects of the ‘social syndrome’ on species-wide genetic diversity, however, depends on a complex combination of N_e , gene flow among populations, and the rates at which populations go extinct and (re)colonize (Harrison & Hastings 1996; Ingvarsson 2002; Pannell & Charlesworth 2000). With limited gene flow, deep divergent lineages harbouring different sets of randomly fixed alleles contribute to maintain species-wide genetic diversity and N_e , even though within-lineage diversity and N_e may be very low. This scenario can be disrupted if population dynamic patterns with high rates of extinction reduce lineage persistence, causing shallow population divergence. Combined with colonization events by some of these less divergent lineages, this type of dynamics may act to eliminate species-wide genetic diversity (Pannell & Charlesworth 2000; Settepani *et al.* 2014; Wade & McCauley 1988). Thereby, metapopulation dynamics may contribute to further elimination of genetic diversity of social inbreeding species.

Interestingly, social inbreeding lineages generally occur at the tip clades of phylogenies indicating their derived state, and showing little evidence for speciation (Agnarsson *et al.* 2006; Chapman *et al.* 2000; Johannesen *et al.* 2007; Jordal & Cognato 2012; Pike *et al.* 2007; Settepani *et al.* 2016). Therefore, while the evolutionary transition from solitary to social living may be favoured by short-term benefits, this social syndrome is not necessarily an evolutionarily stable life form because of the long-term detrimental consequences on genetic diversity and evolutionary potential (Charlesworth 2003; Charlesworth & Charlesworth 1987; Charlesworth & Wright 2001; Romiguier *et al.* 2014a; Wright *et al.* 2013). Stebbins (1957) proposed that evolutionary transitions that reduce N_e and consequently population genetic diversity may lead to an evolutionary dead end, because lineages with low genetic diversity have reduced evolutionary potential and therefore suffer from increased risk of extinction. This hypothesis proposes that species with low genetic diversity are short lived and show reduced diversification rate of lineages (Takebayashi & Morrell 2001). Empirical evidence for the dead end hypothesis comes predominantly from phylogenetic analyses in plants, showing that selfing lineages form short terminal branches and are associated with reduced diversification rates (Charlesworth 2006; Igic *et al.* 2006; Igic & Busch 2013; Takebayashi & Morrell 2001; Wright *et al.* 2013). However, the genetic mechanisms predicted to limit the evolutionary potential of inbreeding lineages such as depletion of genome-wide genetic diversity remain largely untested.

We performed a comprehensive quantitative test of the key prediction that the transition to sociality reduces N_e and genome-wide genetic diversity in individuals and populations, thereby limiting the evolutionary potential in derived social species. Adopting a comparative population genomic approach, we used restriction site associated DNA (RAD) sequencing to generate quantitative estimates of genome-wide genetic diversity in a genus of spiders with contrasting social life forms. The genus *Stegodyphus* contains three independently derived social inbreeding species that each

have a subsocial outcrossing sister species (Johannesen *et al.* 2007; Kraus & Kraus 1988; Settepani *et al.* 2016), providing three independent contrasts for comparing genetic diversity among populations with social inbreeding and subsocial outcrossing mating systems. Our analyses included 235 individuals from multiple populations of these three species pairs and one outgroup, sampled across a wide range of their geographical distribution (Fig. 1). Permanent group living in spiders exists in approximately 25 of more than 46.000 extant spider species (Lubin & Bilde 2007; World Spider Catalog 2017), and remarkably, sociality has evolved independently at least 18 times across 7 families (Agnarsson *et al.* 2006). Sociality in spiders is always associated with elimination of pre-mating dispersal, the transition to an inbreeding mating system (not necessarily associated with a reduction in census population size), female biased sex ratio and reproductive skew (Lubin & Bilde 2007). In contrast, their subsocial sister species maintain pre-mating dispersal and an outcrossing mating system (Bilde *et al.* 2005). In social species, new groups are formed by propagule dispersal of homozygous sib-mated females, and populations are prone to high extinction-colonisation events (Lubin & Bilde 2007). This system is therefore well suited for examining effects of this social life style on population and species-wide genomic diversity.

Materials and methods

Collection of samples

We sampled a total of 232 spiders from seven species of the genus *Stegodyphus* (Eresidae) across multiple geographic localities in Africa, India and Israel. Three of the sampled species are inbreeding and social (*S. mimosarum*, *S. dumicola* and *S. sarasinorum*) and four are outcrossing and subsocial (*S. africanus*, *S. tentoriicola*, *S. pacificus* and *S. lineatus*). We sampled one individual per nest (4-10 nests / locality) from 2-5 localities per species, with an average of nine nests per locality

(sample localities: Fig. 1; see Table S1 for detailed sampling information). A nest refers to an adult female with her brood for subsocial species and a permanent group of adults and their offspring for social species. We refer to individuals from the same locality as a population. Samples of *S. sarasinorum* were collected in October-December 2010 and January 2012, *S. pacificus* in January 2012, *S. lineatus*, *S. mimosarum*, *S. africanus* and *S. dumicola* were collected in April-June 2012 and *S. tentoriicola* in May 2012 and November 2013.

Library construction, sequencing and data processing

Genomic DNA was extracted using the DNeasy Blood & Tissue kit, QIAGEN (Germantown, MD, USA). DNA from each individual was normalized to 20 ng/μl in 10 μl and double-digested with *SbfI* and *MspI* (New England Biolabs). Paired-end RAD tag libraries were created following the protocol in Poland *et al.* (2012) with the following modifications: 0.5 μl of BSA was added to the Restriction Mastermix and a final AMPure Beads clean-up and size selection step was performed after PCR amplification of the libraries to obtain a size range of approximately 150-600 bp. Each individual sample was barcoded with a unique sequence ranging from 4 nt to 9 nt in length (Poland *et al.* 2012). To obtain paired end reads (100 bp reads), pools of 40 individuals were sequenced on a single Illumina sequencing lane. The pipeline for data cleaning and processing is described in Fig. S1.

Proportion of protein coding sequence in Stegodyphus

A reference genome of the social inbreeding *Stegodyphus mimosarum* (Sanggaard *et al.* 2014), was estimated to be 2.7 gb long, and to carry about 27,000 genes with an average length of about 1,000 bp, suggesting that only approximately 1 % of the genome is protein coding. To estimate the proportion of protein coding RAD loci, we mapped the RAD loci obtained from *S. mimosarum* to its reference genome. As expected, our calculations show that only 1.16 % of the RAD loci are

protein coding. We assume that a similarly low proportion of RAD loci are protein coding in the other *Stegodyphus* species, and that our data therefore mainly represent non-coding DNA sequences.

Data analyses

The cleaned and processed data were analysed using pyRAD_v.2.15 (within-sample clustering, error-rate and heterozygosity estimation, creation of consensus sequences, clustering of consensus sequences across samples in a population, alignment and detection of paralogs) (Eaton 2014).

pyRAD analyses were performed using the following parameters: ‘Minimum coverage for a cluster’ 10; ‘Clustering threshold’ 0.95; ‘Minimum samples in a final locus’ was adjusted according to the number of individuals in the given population; ‘Maximum number of individuals with shared heterozygote sites’ 60 %; ‘Ns in a consensus sequence’ 110; ‘Heterozygote sites in a consensus sequence’ 3; ‘Max number of SNPs in a final locus’ 10; ‘Max stack size’ 30; ‘Hierarchical cluster groups’ were specified for each species according to populations. For all analyses, indels and their flanking regions (7 bp) were masked to avoid misalignment.

We used custom scripts to further process the output of the pyRAD analyses in order to estimate genetic diversity (π) and genetic structure. To perform population genetic analyses it was necessary to remove any bias from the difference in RAD loci length and the potential for lack of independence between RAD loci. We therefore concatenated accepted loci into one alignment in a random order before analysis. For all diversity estimates we split each sample into 20 subsets of equal size, and estimated the population genetic parameter on each subset to obtain confidence intervals. All confidence intervals were obtained by bootstrapping and were used for statistical comparisons. To ensure that each population was represented equally within a species, populations were down-sampled to the number of individuals present in the population with the lowest number of individuals. The minimum number of individuals required in all populations in order to keep a

locus was set to five. However, due to low numbers of individuals in the subsocial *S. pacificus* (four individuals per population) and the social *S. sarasinorum* (seven individuals in population D), analyses of these species were performed with two and four as a minimum number of individuals needed in all populations to keep the loci, respectively.

We could not use the *S. mimosarum* reference genome to filter RAD sites as it could only be used for *S. mimosarum* and partially for its sister species *S. africanus*. The remaining species were too divergent and only ~1 % of the reads mapped to the *S. mimosarum* reference genome with high certainty. We decided to do all analyses de novo not to add any bias.

We used MEGA6 (Tamura *et al.* 2013) to estimate genetic diversity (π and θ) and Tajima's D per population and per species. Genetic differentiation (F_{st}) was estimated by $1 - \pi_P/\pi_S$ with π_P being the average π of the populations under consideration, and π_S the total π of the species under consideration. Population structure analyses were performed in InStruct (Gao *et al.* 2007) in Mode 4 i.e. to infer population structure with admixture and population inbreeding coefficients, for $K = 1$ to $K = N + 1$, K being the number of genetic clusters and N being the number of geographical populations in the species under consideration, and fastSTRUCTURE (Raj *et al.* 2014) for $K = 1$ to $K = N + 1$. Structure analyses were performed on input files containing the informative sites of the given sample. Graphical representation of structure results were obtained with a custom script (available here: <https://github.com/shenglin-liu/plotFastStructure>) for fastStructure and distruct1.1 (Rosenberg 2004) for InStruct.

Individual heterozygosity (H_{obs}) based on all sites with coverage of at least 10 were estimated as the number of heterozygote sites divided by the total number of sites. To test the difference in H_{obs} between subsocial and social sister species we constructed a general linear mixed model in the R package "lme4" (Bates *et al.* 2015) for each species pair. Each model contained mating system (subsocial or social) as the sole fixed effect and population as a random effect. P-values were

obtained by comparing a reduced model, without mating system, with the full model using a likelihood ratio test. H_{obs} was square-root transformed to fulfil assumptions of parametric analysis. We used a two-sided Mann-Whitney U test to compare differences between F_{st} and π values between social and subsocial species. Analyses were performed in R (R Development Core Team 2016).

A summary of the characteristics of the data sets used, i.e. number of individuals and populations for each species, number of sites and polymorphic sites, total number of reads and loci, are reported in Table S2.

Demographic modelling

Coalescent simulations were performed to estimate demographic parameters using fastsimcoal2 (Excoffier *et al.* 2013). First, for each RAD locus included in the diversity analyses (see criteria above), the number of sequences per population were randomly down-sampled to the same number for all RAD loci. The sequence number depended on the number of samples per population for the given species and the amount of data retained per individual, and thereby presents a trade-off between number of sequences and amount of data per individual. Secondly, folded joint site frequency spectra (SFSs) were calculated from the combination of all RAD loci using a custom R script (available here: <https://github.com/shenglin-liu/vcf2sfs>) (for graphical representation see Fig. S2). Population models were constructed based on pairwise population divergence times, estimated also using fastsimcoal2. It was not possible to reconstruct a bifurcating population model for *S. tentoriicola*, so we used a star model. One hundred independent fastsimcoal2 runs were performed with 500,000 coalescent simulations, as well as 10–40 cycles of the likelihood maximization algorithm. We used a mutation rate of 6.65E-9, which is the average of the two differently estimated mutation rates used by Mattila *et al.* (2012) and a generation time of one year as *Stegodyphus* species are annual (Johannesen *et al.* 2007). We estimated separate population sizes on

each branch (assuming constant population sizes), and all population split times. For each species, we also constructed a model including migration among all current populations, with separate migration rates between all possible population pairs. Confidence limits were estimated the following way: all RAD loci were concatenated, and 100 new data sets were constructed by bootstrapping over sites using seqboot from the PHYLIP package (Felsenstein 1989). For each new data set, joint frequency spectra were calculated as described above and 50 independent fastsimcoal2 runs were performed for each bootstrapped data set. By comparing maximum likelihood values between runs, we obtained the best run from each dataset, and calculated confidence limits for the model based on 100 sets of parameters.

Results

Genetic diversity

We estimated genetic diversity at three levels: Observed individual heterozygosity (H_{obs}), within population diversity (π_s), and species-wide diversity (π_T). Observed individual heterozygosity (H_{obs}) estimates were on average more than six times lower in the social inbreeding than the subsocial outcrossing species (*S. pacificus* – *S. sarasinorum*: $\chi^2 = 13.47$, $df = 1$, $P = 0.0003$; *S. tentoriicola* – *S. dumicola*: $\chi^2 = 51.05$, $df = 1$, $P < 0.0001$; *S. africanus* – *S. mimosarum*: $\chi^2 = 113.73$, $df = 1$, $P < 0.0001$) (Fig. 2A). Average H_{obs} of the social species was estimated to 0.00041, ranging from 0.00032 (*S. mimosarum*) to 0.00050 (*S. sarasinorum*), while average H_{obs} of the subsocial species was estimated to 0.00265, ranging from 0.00210 (*S. africanus*) to 0.00301 (*S. lineatus*) (Fig. 2A). Average within population nucleotide diversity (π_s) was estimated to be more than ten times lower in the social inbreeding species ($\pi_s = 0.00031$) compared with subsocial outcrossing species ($\pi_s =$

0.00393) (Fig. 2B, Table S3). The lowest social population diversity was estimated in the *S. dunicola* population from Paarl (PAA) with $\pi_s = 0.00003$, while the highest diversity estimate was from the southern *S. sarasinorum* population (C), $\pi_s = 0.00077$. The diversity of subsocial populations ranged from $\pi_s = 0.00257$ in *S. africanus* (PON) to $\pi_s = 0.00834$ in *S. pacificus* (R2) (Fig. 2B, Table S3). The contrast in species-level genetic diversity among inbreeding and outcrossing species may be affected by differences in geographic distances among sampling sites. Since in our case, outcrossing populations in some instances were sampled at smaller ranges than inbreeding species (Fig. 1), our results should be robust, because similar distances among sampling sites is expected to translate into even more pronounced differences in the observed diversity patterns.

Species-wide nucleotide diversities (π_T) were on average more than four times lower in the inbreeding compared to outcrossing species (Fig. 2B, Table S3). Species-wide nucleotide diversity (π_T) estimate for the social *S. mimosarum* was higher than for the two other social species. This is due to the presence of two differentiated lineages of populations (the Malagasy and South African cluster) that are estimated to have diverged about 250,000 years ago, and are most likely completely isolated from each other by the Mozambique Chanel. When considering these two genetic lineages independently, estimates were similar for all social lineages and show that species-wide genetic diversity was eight times lower for social inbreeding species than for subsocial outcrossing species ($\pi_T = 0.00056$ vs. $\pi_T = 0.00474$) (*S. pacificus* - *S. sarasinorum*: $U = 0$, $n1 = n2 = 19$, $P < 0.001$; *S. tentoriicola* - *S. dunicola*: $U = 0$, $n1 = 18$, $n2 = 19$, $P < 0.001$; *S. africanus* - *S. mimosarum*: $U = 0$, $n1 = n2 = 19$, $P < 0.0001$; *S. africanus* - *S. mimosarum* Madagascar: $U = 0$, $n1 = n2 = 19$, $P < 0.0001$; *S. africanus* - *S. mimosarum* South Africa: $U = 0$, $n1 = n2 = 19$, $P < 0.0001$) (Fig. 2B). Values of θ and Tajima's D for each population and species are included in Fig. S3. The θ values are qualitatively very similar to the π estimates with lower values in the social inbreeding compared

to the subsocial outcrossing species. All Tajima's D estimates were negative with no consistent differences between the social and subsocial species.

Genetic structure

The InStruct and fastSTRUCTURE analyses showed that a single genetic cluster is the most likely in each of the four subsocial outcrossing species ($K = 1$), while more than one genetic cluster is most likely in each of the three social inbreeding species (model selection was based on Deviance Information Criteria) (Fig. 3A InStruct, Fig. S4 fastSTRUCTURE). The inferred genetic clusters for the social species corresponded closely to predefined populations. The individuals from two *S. sarasinorum* populations (A and B) formed one genetic cluster, as did the ADDO and PAA, and PON and KRU populations of *S. dumicola*. Two genetic clusters were the most likely when considering all *S. mimosarum* populations, however, when performing the analyses on the Malagasy and South African populations separately, each predefined population constituted its own genetic cluster (Fig. 3A, Fig. S4). See Fig. S4 for a representation of the best fastSTRUCTURE estimates of K per species. InStruct simultaneously estimates inbreeding coefficients (F_{IS}) for the genetic clusters. Inbreeding coefficients for social species were substantially higher than those of subsocial species (Fig. 3). *Stegodyphus sarasinorum* populations showed the lowest inbreeding coefficients of the social species ranging from 0.55 to 0.71, while *S. dumicola* populations have the highest ranging from 0.71 to 0.88. Inbreeding coefficients of subsocial species were 0.15-0.17 except for *S. pacificus* with 0.44. We only have a few samples from this species and attribute the relatively high inbreeding coefficient to stochastic effects.

Differentiation among populations measured by F_{ST} was more than 10 times lower in subsocial outcrossing compared to social inbreeding species (*S. pacificus* - *S. sarasinorum*: $U = 361$, $n_1 = n_2 = 19$, $P < 0.001$; *S. tentoriicola* - *S. dumicola*: $U = 361$, $n_1 = n_2 = 19$, $P < 0.001$; *S. africanus* - *S.*

mimosarum: $U = 361$, $n1 = n2 = 19$, $P < 0.0001$; *S. africanus* - *S. mimosarum* Madagascar: $U = 361$, $n1 = n2 = 19$, $P < 0.0001$; *S. africanus* - *S. mimosarum* South Africa: $U = 361$, $n1 = n2 = 19$, $P < 0.0001$) (Fig. 3B). Average F_{st} of subsocial and social species was 0.046 and 0.641, respectively. The estimate for social species is strongly inflated by the physical separation of the Malagasy and South African populations; however, a tenfold difference in F_{st} between the subsocial and social species remained when considering the Malagasy and South African populations separately (average social $F_{st} = 0.467$) (Fig. 3B).

Demographic modelling

We found a tenfold reduction in N_e of social inbreeding species compared to subsocial outcrossing species (population level estimates, Fig. 4). Also, the time to the most recent common ancestor (TMRCA) was considerably longer for subsocial species compared to social species (Fig. 4). We note that the estimates of N_e and TMRCA especially in the subsocial species appear to deviate from coalescent based predictions (either too high N_e or too low TMRCA), it is possible that that this deviation is explained by migration. *Stegodyphus mimosarum* was an exception, however, but it is reasonable to assume two independent population dynamic processes due to the deep split between populations from Madagascar and South Africa caused by the physical barrier created by Mozambique Channel. Average migration rates were estimated to be lower in social relative to subsocial species, as expected given inbreeding in social species (approximately by a factor two). Models including migration provide the best fit for all species except for *S. pacificus*, based on Akaike Information Criterion. We note that the estimated parameters of demographic modelling may be biased due to selection on the markers used or on linked loci (Schridder *et al.* 2016). The extent of such bias depends on the combination of the strength and efficacy of selection (effective population size and the selection coefficient, $N_e s$) and the extent of linkage disequilibrium. Since $N_e s$ is predicted to be higher in the subsocial species, while the extent of linkage disequilibrium is

predicted to be higher in the social species, it is difficult to predict if the estimated parameters are biased more in social than subsocial species.

Discussion

By contrasting three independent lineages of social inbreeding species with their subsocial outcrossing sister species, we found extraordinary low estimates of genetic diversity for social species at all levels. Estimates of population level neutral genetic diversity for a wide range of species (Leffler *et al.* 2012) show that mammals such as the European lynx (*Lynx lynx*) with extremely small small breeding populations (as low as 40-50 individuals are present in the Balkan region (IUCN 2015)) have extremely low population genetic diversity ($\sim \pi = 0.0001$). Social spiders showed up to three times lower population genetic diversity compared with the lynx (*S. dumicola* in Paarl, $\pi_S = 0.00003$), despite the fact that there are many thousand individuals of *S. dumicola* in the population (Bilde *et al.* 2007; Lubin & Bilde 2007). This suggests that factors associated with the transition to a social life style including female biased sex ratio, reproductive skew, meta-population dynamics and inbreeding, in addition to breeding population size significantly affect individual, population-level and species-wide genetic diversity. While we cannot disentangle the effects of the different characteristics of the social lifestyle, our results unequivocally demonstrate that the evolution of this social syndrome incurs dramatic loss of genome-wide genetic diversity at all levels.

The evolutionary transition to sociality with repeated cycles of inbreeding and the propagation of lineages by single mated females is expected to strongly decrease individual heterozygosity. In accordance, we found that estimates of observed individual heterozygosity (H_{obs}) were much lower in the three social species compared to their outcrossing congeners (Fig. 2A). At the population level, theory predicts that inbreeding alone can reduce genetic diversity by at most 50% (self-

fertilization) (Charlesworth 2003; Charlesworth & Wright 2001; Nordborg 2000). However, an even stronger reduction of up to $\sim 2/3$ is found in some selfing plant species (Foxy *et al.* 2009). This is attributed to lower effective recombination rates and stronger effects of hitchhiking events in selfing species. We document an even more severe reduction in genetic diversity of social species as estimates of population nucleotide diversity ($\pi_s < 0.0001$) was 6-fold lower in social populations compared to subsocial populations (Fig. 2B). The higher reduction in population genetic diversity than expected solely through inbreeding may be caused by hitchhiking events. Linkage disequilibrium is predicted to be extended due to reduced effective recombination rate in the social inbreeding species (Charlesworth & Wright 2001), and hitchhiking events are therefore expected to remove more linked variation. However, N_e is lower in the social species than in subsocial species (Fig. 4) leading to lower efficacy of selection (Settepani *et al.* 2016), which potentially reduces the number of hitchhiking events (Charlesworth & Wright 2001). In addition, preliminary evidence from humans and *Drosophila* suggests that most selective sweeps are so-called ‘soft’ sweeps fixing existing genetic variants (Cutter & Payseur 2013). Because subsocial species harbour more standing genetic variation, loss of diversity by selective sweeps is predicted to be higher in subsocial outcrossing species than social inbreeding species. While we cannot rule out that hitchhiking may have contributed to the observed loss of genetic diversity, we suggest that female biased sex ratio, reproductive skew and potentially metapopulation dynamics (see discussion below) are the main drivers of the observed loss in genetic diversity. These factors are inherent to the social syndrome and act to further reduce N_e relative to the expected reduction in N_e to 50% by inbreeding alone (Wright 1931).

Furthermore, metapopulation dynamics with high nest and population turnover (Ingvarsson 2002; Pannell & Charlesworth 2000), also contribute to the depletion of genetic diversity in the social

species. We will discuss metapopulation processes further in relation to species-wide genetic diversity below.

In the absence of gene flow, social lineages should become fixed for different alleles through neutral or adaptive divergence, which means that genetic diversity may be maintained at the species level. In contrast with this expectation, species-wide genetic diversity (π_T) of social spiders was 5-8 fold lower than that of subsocial species, and amongst the lowest estimated for any species (Leffler *et al.* 2012). One of the social species, *S. mimosarum*, showed higher species-wide genetic diversity compared to the two other social species. However, population structure analyses of *S. mimosarum* revealed two genetic lineages, one in South Africa and one in Madagascar, that diverged approximately 250,000 years ago (Fig. 4). Genetic diversity within each of these lineages was comparable to the two other social species. If we consider the Malagasy and South African lineages separately, which is justified by the separation caused by ~1000 km ocean and genetic divergence through isolation by distance, we observe similar coalescence times of all social inbreeding species (Fig. 4). Under this separation of *S. mimosarum* into two independent genetic lineages, species-wide diversity estimates in all the inbreeding *Stegodyphus* are lower than those reported in other species with inbreeding mating systems (Graustein *et al.* 2002; Walser & Haag 2012). It is possible that the difference in species-level genetic diversity among social and subsocial species could be affected by differences in geographic distances among sampling sites. However, sampling of subsocial populations were always performed at similar or smaller ranges than sampling of social species (Fig. 1). We therefore consider our results robust, because similar distances among sampling sites is expected to translate into even more pronounced differences in the observed diversity patterns.

The existence of polymorphisms within the restriction site when using RAD sequencing may result in an underestimation of genetic diversity (Arnold *et al.* 2013). Our results are very robust to this inherent bias, because higher diversity, as observed in the subsocial species, will result in more

polymorphic restriction sites and lead to an underestimation of genetic diversity compared to the social species. For diversity data similar to those observed/estimated in our study, underestimation of genetic diversity is expected to be in the order of 10-15 % (Arnold *et al.* 2013), and correcting for this bias would not change our the result of extremely low estimates of genetic diversity in the social species.

In line with theoretical expectations, demographic modelling consistently produced tenfold lower N_e estimates in the three social species compared to their subsocial sister species (Fig. 4). The combined effect of low N_e and restricted gene flow due to lack of pre-mating dispersal and inbreeding predicts strong differentiation among populations (Harrison & Hastings 1996; Ingvarsson 2002; Pannell & Charlesworth 2000). Our results partly corroborate this prediction as genetic and geographic clusters closely matched and were highly differentiated (Fig. 3). This could possibly be explained by relatively longer average geographic distance between sampled populations of the social species. On the other hand, the distances between the outcrossing *S. tentoriicola* and *S. africanus* populations are comparable to the distances between both the South African and the Malagasy social *S. mimosarum* populations, yet *S. tentoriicola* form a single genetic cluster, while the *S. mimosarum* populations form separate clusters. The divergence among social populations is, however, based on relatively few fixed SNPs. In the absence of gene flow, population divergence is expected to increase over time, contributing to the maintenance of genetic diversity at the species level (Charlesworth 2003), but this is not what we found in the social species. Despite strong differentiation between populations, the divergence between populations was shallow and the time to the most recent common ancestor in the social species was shorter (on average less than half of that of subsocial species) (Fig. 4). The extent of divergence depends on population age, population turnover rates, and extent of gene flow among populations (Barton & Whitlock 1997; Wade & McCauley 1988; Whitlock & McCauley 1990).

Population turnover is expected to deplete within-population diversity due to genetic bottlenecks/founder events (Ingvarsson 2002), but to increase between-population differentiation due to genetic drift and strong founder effects (Harrison & Hastings 1996; Pannell & Charlesworth 2000) e.g. in aphids (Massonnet *et al.* 2002), or *Daphnia* and other freshwater invertebrates (Freeland *et al.* 2000; Walser & Haag 2012). Our data are consistent with a scenario where social populations differentiate, as predicted by theory, but divergence is counteracted by high turnover rates of genetic lineages (Settepani *et al.* 2014). If genetic lineages were recently established because turnover rate is particularly high (Wade & McCauley 1988), we would expect to observe genetic homogenization and therefore low between-population divergence and low species-wide genetic diversity (Haag *et al.* 2005). Indeed, field studies show that social spider populations are characterised by rapid nest turnover and high population extinction (Bilde *et al.* 2007; Crouch & Lubin 2001), which is consistent with this scenario.

Genetic homogenization by lineage extinction-colonization dynamics requires social spiders to be able to disperse over large distances and successfully establish new populations. Lineages are propagated by mated females (Johannesen *et al.* 2002), and long-distance dispersal occurs by ballooning, enabling dispersal over much longer distances than simply by walking (Schneider *et al.* 2001). Propagule dispersal is expected to occur when nests overshoot their optimal size (Bilde *et al.* 2007; Ulbrich & Henschel 1999), or when local conditions deteriorate (Crouch & Lubin 2001). Random dispersal by ballooning is risky, and the likelihood of successfully establishing a new nest is low as small nests experience high mortality (Bilde *et al.* 2007; Johannesen *et al.* 2002). The population dynamic processes associated with lineage propagation therefore act to homogenize genetic variation among populations (Settepani *et al.* 2014), which may explain the extremely low species-wide diversity in social species. In relation to this scenario, we note that both structure

analyses and demographic models are most consistent with the occurrence of some degree of gene flow. However, even very low occurrence of ballooning individuals entering and joining already existing populations or nests would account for these results. This does not necessarily mean that individuals originating from different populations admix; they may co-exist as independent lineages in a location (Johannesen *et al.* 2002), until one of the lineages die out. Indeed, population structure analysis provides only limited evidence of admixture of genotypes originating from different genetic lineages (*S. sarasinorum*, population C and D, Fig. 3).

Consequences of low genetic diversity

The extent and spatial subdivision of genetic diversity has important consequences for individual fitness and population viability (Hughes *et al.* 2008; Reed & Frankham 2003; Spielman *et al.* 2004b; Whitehorn *et al.* 2011), and for evolutionary processes such as local adaptation and speciation (Pannell & Charlesworth 2000; Wright 1931). Our data show that social species with female bias, reproductive skew and inbreeding show extremely low population and species-wide genetic diversity, which is likely to limit their ability to respond to environmental or ecological challenges (Fox & Reed 2011; Schou *et al.* 2015). Indeed, phylogenetic patterns of taxa that show this social syndrome indicate that the social lineages are derived, short-lived, and appear to lack diversification (Agnarsson *et al.* 2006; Chapman *et al.* 2000; Johannesen *et al.* 2007; Jordal & Cognato 2012; Pike *et al.* 2007; Settepani *et al.* 2016). In addition, low N_e is expected to reduce effective recombination rate and efficacy of selection, which may lead to a build-up of the genetic load by the accumulation of deleterious mutations (Charlesworth 2003; Charlesworth & Wright 2001). Evidence for reduced efficacy of selection associated with mating system transitions was found in several systems as for example in plants (Arunkumar *et al.* 2015), freshwater snails (Burgarella *et al.* 2015) and eusocial insects (Romiguier *et al.* 2014b), in addition to social spiders

(Mattila *et al* 2012; Settepani *et al.* 2016). Deleterious effects of low N_e may result in a less effective immune response towards pathogens (Cassinello *et al.* 2001; Spielman *et al.* 2004a). In social *Stegodyphus* spiders, it is hypothesized that high extinction rates of nests and entire populations may be caused by elevated vulnerability to diseases (Bilde *et al.* 2007; Crouch & Lubin 2001), and preliminary data indicated that social *Stegodyphus* species have a less effective immune response compared to outcrossing species (Jönsson 2015). Deleterious effects of inbreeding on the immune system may cause a decline in population fitness (Acevedo-Whitehouse *et al.* 2003; Coltman *et al.* 1999; Liersch & Schmid-Hempel 1998), and contribute to increase risk of extinction of populations with small effective size.

The evolution of group living and cooperation is widespread in the animal kingdom, and is attributed an important role for the ecological success of dominant species such as humans (Fehr & Fischbacher 2003; Kramer 2010) and social insects (Wilson 1971). Most social birds, mammals and insects maintain pre-mating dispersal or other inbreeding avoidance mechanisms and outcrossing mating systems (Pusey & Wolf 1996; Pusey 1987; Tabadkani *et al.* 2012). Thereby, they avoid the immediate fitness cost caused by the expression of deleterious alleles (Charlesworth & Willis 2009; Wang *et al.* 1999), which is expected under inbreeding even if some of the genetic load is purged by natural selection (Crnokrak & Barrett 2002; Hedrick 1994). Furthermore, because of outcrossing, N_e is higher compared to inbreeding social species and the long term costs in the form of reduced efficacy of selection and accumulation of deleterious mutations is therefore relatively lower (see references above). Nevertheless, the evolution of sociality is tightly connected with factors that diminishes N_e also in outcrossing species, most prominently reproductive skew, but also restricted dispersal (Cockburn 2003). Genetic diversity (π_s) of RAD loci in bumblebees (Lozier 2014) and of transcriptomes of eusocial termites, ants and bees (Romiguier *et al.* 2014a) was shown to be similar

to that of subsocial spider species in this study, and although their diversity is thereby approximately 10x higher than that of the social inbreeding spider species, it is still low compared with non-social outcrossing systems (Leffler 2012 Romiguier *et al.* 2014a). Within social insects, the level of social complexity across ant species correlates positively with reduced N_e , reduced genetic diversity, and increased genetic load (Romiguier *et al.* 2014b), suggesting that despite outcrossing, effects of high reproductive skew and demography have detrimental consequences for genomic architecture and population genetic diversity. In the social taxa that are characterized by lack of outcrossing (Agnarsson *et al.* 2006; Chapman *et al.* 2000; Johannesen *et al.* 2007; Jordal & Cognato 2012; Pike *et al.* 2007; Settepani *et al.* 2016), the negative effects of low N_e on population genetic diversity and build-up of the genetic load are even further amplified by inbreeding. The short-term benefits of sociality in these taxa therefore come with long-term costs in the form of reduced adaptive potential, consistent with the hypothesis that this social syndrome may be an evolutionary dead end.

Acknowledgements

This work was supported by the European Research Council (ERC StG-2011_282163 to TB) and Augustinus Fonden (<http://www.augustinusfonden.dk>) supported VS for expenses related to sampling in South Africa and Madagascar. We acknowledge the provincial agencies involved in the issuing of permits in South Africa: Eastern Cape – Permits nos. CRO 63/12CR and CRO 64/12 CR; KwaZulu Natal – Permit no. OP 1896/2012; Limpopo Province – Permit no. 001-CPM402-00001, South African National Parks – Permit no. BILDT1008. We thank Susanna Clusella-Trullas for providing the spiders collected under Cape Nature permit 0035-AAA007-00428, and Jonathan Pruitt for the collection permits in the Northern Cape Province – Permit no. FAUNA 189/2013. We

thank Christina Holm for collection of the *S. lineatus* samples, Rohini Balakrishnan, Agastya International Foundation and Ms Rao for the support provided in India, and MICET (Malagasy association for the Conservation of Tropical Ecosystems) for technical guidance for collection permits and transportation in Madagascar. We are also indebted to Jinjie Duan, Mikkel Schierup, Michael M. Hansen and Magnus W. Jacobsen for useful discussion and suggestions, Linda Hartun Konggaard for the guidance and great help in the preparation of the reduced representation libraries, Shenglin Liu for sharing the R script used to create structure graphs, Michael Glad for technical support and Deren Eaton for his prompt and kind help with pyRAD.

Data accessibility.

Raw sequence data for each individual is available under Accession no. SRP095316 at the Sequence Read Archive (SRA) at NCBI (<http://www.ncbi.nlm.nih.gov/sra>), and the associated BioProject is PRJNA356501. Sequence alignment data is available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.hg785>

Author contribution.

Conceived and designed the experiments: TB JB. Collected the samples: VS MG LG. Performed the laboratory work: VS. Analysed the data: VS JB MFS. Wrote the paper: VS MFS JB TB LG MG.

Conflict of Interest.

The authors declare no competing financial interests.

References

- Acevedo-Whitehouse K, Gulland F, Greig D, Amos W (2003) Disease susceptibility in California sea lions. *Nature* **422**, 35-35.
- Agnarsson I, Aviles L, Coddington JA, Maddison WP (2006) Sociality in Theridiid spiders: Repeated origins of an evolutionary dead end. *Evolution* **60**, 2342-2351.
- Alcala N, Vuilleumier S (2014) Turnover and accumulation of genetic diversity across large time-scale cycles of isolation and connection of populations. *Proceedings of the Royal Society B-Biological Sciences* **281**.
- Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology* **22**, 3179-3190.
- Arunkumar R, Ness RW, Wright SI, Barrett SCH (2015) The Evolution of Selfing Is Accompanied by Reduced Efficacy of Selection and Purging of Deleterious Mutations. *Genetics* **199**, 817-+.
- Barton NH, Whitlock MC (1997) The Evolution of Metapopulations. In: *Metapopulation Biology*, pp. 183-210. Academic Press, San Diego.
- Bates D, Machler M, Bolker BM, Walker SC (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1-48.
- Bilde T, Coates KS, Birkhofer K, et al. (2007) Survival benefits select for group living in a social spider despite reproductive costs. *Journal of Evolutionary Biology* **20**, 2412-2426.
- Bilde T, Lubin Y (2011) Group living in spiders: cooperative breeding and coloniality. In: *Spider Behaviour - flexibility and versatility* (ed. Herberstein M). Cambridge University Press, Cambridge.
- Bilde T, Lubin Y, Smith D, Schneider JM, Maklakov AA (2005) The transition to social inbred mating systems in spiders: role of inbreeding tolerance in a subsocial predecessor. *Evolution* **59**, 160-174.
- Bourke AFG (1997) Sex ratio in bumblebees. *Philosophical transactions of the royal society B: Biological sciences* **352**, 1921-1933.
- Burgarella C, Gayral P, Ballenghien M, et al. (2015) Molecular Evolution of Freshwater Snails with Contrasting Mating Systems. *Molecular Biology and Evolution* **32**, 2403-2416.
- Cassinello J, Gomendio M, Roldan ERS (2001) Relationship between coefficient of inbreeding and parasite burden in endangered gazelles. *Conservation Biology* **15**, 1171-1174.
- Chapman TW, Crespi BJ, Kranz BD, Schwarz MP (2000) High relatedness and inbreeding at the origin of eusociality in gall-inducing thrips. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 1648-1650.
- Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**, 195-205.
- Charlesworth D (2003) Effects of inbreeding on the genetic diversity of population. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **358**, 1051-1070.
- Charlesworth D (2006) Evolution of plant breeding systems. *Current Biology* **16**, R726-R735.
- Charlesworth D, Charlesworth B (1987) Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* **18**, 237-268.
- Charlesworth D, Willis JH (2009) Fundamental concepts in genetics - The genetics of inbreeding depression. *Nature Reviews: Genetics* **10**, 783-796.
- Charlesworth D, Wright SI (2001) Breeding systems and genome evolution. *Current Opinion in Genetics & Development* **11**, 685-690.
- Clobert J, Danchin E, Dhondt AA, Nichols JD (2001) *Dispersal* Oxford University Press, Oxford.
- Cockburn A (1998) Evolution of helping behavior in cooperatively breeding birds. *Annual Review of Ecology and Systematics* **29**, 141-177.
- Cockburn A (2003) Cooperative breeding in oscine passerines: does sociality inhibit speciation? *Proceedings of the Royal Society B-Biological Sciences* **270**, 2207-2214.

- Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* **53**, 1259-1267.
- Crnokrak P, Barrett SCH (2002) Perspective: Purging the genetic load: A review of the experimental evidence. *Evolution* **56**, 2347-2358.
- Crouch T, Lubin Y (2001) Population stability and extinction in a social spider *Stegodyphus mimosarum* (Araneae : Eresidae). *Biological Journal of the Linnean Society* **72**, 409-417.
- Cutter AD, Payseur BA (2013) Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics* **14**, 262-274.
- Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* **30**, 1844-1849.
- Ellegren H, Galtier N (2016) Determinants of genetic diversity. *Nature Reviews Genetics* **17**, 422-433.
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data *PLoS Genetics* **9(10)**, e1003905.
- Fehr E, Fischbacher U (2003) The nature of human altruism. *Nature* **425**, 785-791.
- Felsenstein J (1989) PHYLIP -- Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164-166.
- Fox CW, Reed DH (2011) Inbreeding depression increases with environmental stress: an experimental study and meta-analysis. *Evolution* **65**, 246-258.
- Foxe JP, Slotte T, Stahl EA, *et al.* (2009) Recent speciation associated with the evolution of selfing in *Capsella*. *Proceedings of the National Academy of Sciences, United States of America* **106**, 5241–5524.
- Freeland, Noble, Okamura (2000) Genetic consequences of the metapopulation biology of a facultatively sexual freshwater invertebrate. *Journal of Evolutionary Biology* **13**, 383-395.
- Gao H, Williamson S, Bustamante CD (2007) A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* **176**, 1635-1651.
- Graustein A, Gaspar JM, Walters JR, Palopoli MF (2002) Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. *Genetics* **161**, 99-107.
- Haag CR, Riek M, Hottinger JW, Pajunen VI, Ebert D (2005) Genetic diversity and genetic differentiation in *Daphnia* metapopulations with subpopulations of known age. *Genetics Society of America* **170**, 1809-1820.
- Hamilton WD (1967) Extraordinary Sex Ratios. *Science* **156**, 477-488.
- Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **351**, 1291-1298.
- Harrison S, Hastings A (1996) Genetic and evolutionary consequences of metapopulation structure. *Trends in Ecology & Evolution* **11**, 180-183.
- Hatchwell BJ, Komdeur J (2000) Ecological constraints, life history traits and the evolution of cooperative breeding. *Animal Behaviour* **59**, 1079-1086.
- Hedrick PW (1994) Purging Inbreeding Depression and the Probability of Extinction - Full-Sib Mating. *Heredity* **73**, 363-372.
- Hedrick PW, Parker JD (1997) Evolutionary genetics and genetic variation of haplodiploids and X-linked genes. *Annual Review of Ecology and Systematics* **28**, 55-83.
- Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M (2008) Ecological consequences of genetic diversity. *Ecology Letters* **11**, 609-623.
- Igic B, Bohs L, Kohn JR (2006) Ancient polymorphism reveals unidirectional breeding system shifts. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 1359-1363.
- Igic B, Busch JW (2013) Is self-fertilization an evolutionary dead end? *New Phytologist* **198**, 386-397.
- Ingvarsson PK (2002) A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. *Evolution* **56**, 2368-2373.
- IUCN (2015) *The IUCN red list of threatened species*. <http://www.iucnredlist.org/details/12519/0>

- Johannesen J, Hennig A, Dommermuth B, Schneider JM (2002) Mitochondrial DNA distributions indicate colony propagation by single matri-lineages in the social spider *Stegodyphus dumicola* (Eresidae). *Biological Journal of the Linnean Society* **76**, 591-600.
- Johannesen J, Lubin Y, Smith DR, Bilde T, Schneider JM (2007) The age and evolution of sociality in *Stegodyphus* spiders: a molecular phylogenetic perspective. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **274**, 231-237.
- Johnson PCD, Llewellyn KS, Amos W (2000) Microsatellite loci for studying clonal mixing, population structure and inbreeding in a social aphid, *Pemphigus spyrothecae* (Hemiptera : Pemphigidae). *Molecular Ecology* **9**, 1445-1446.
- Jønsson AK (2015) Comparative study of hemolymph antimicrobial activity Indicates a lower immune efficiency of social inbreeding *Stegodyphus* spiders compared to their sub-social outcrossing sister species. *Master thesis*.
- Jordal BH, Cognato AI (2012) Molecular phylogeny of bark and ambrosia beetles reveals multiple origins of fungus farming during periods of global warming. *Bmc Evolutionary Biology* **12**, 133.
- Keller L, Peer K, Bernasconi C, Taborsky M, Shuker DM (2011) Inbreeding and selection on sex ratio in the bark beetle *Xylosandrus germanus*. *Bmc Evolutionary Biology* **11**, 359.
- Koenig WD, Pitelka FA, Carmen WJ, Mumme RL, Stanback MT (1992) The evolution of delayed dispersal in cooperative breeders. *Quarterly Review of Biology* **67**, 111-150.
- Kramer KL (2010) Cooperative Breeding and its Significance to the Demographic Success of Humans. *Annual Review of Anthropology, Volume 39* **39**, 417-436.
- Kraus O, Kraus M (1988) The genus *Stegodyphus* (Arachnida, Araneae). Sibling species, species groups, and parallel origin of social living. In: *Verhandlungen des Naturwissenschaftlichen Vereins in Hamburg (NF) 30* (ed. Kraus O), pp. 151-254. Verlag Paul Parey, Hamburg and Berlin.
- Leffler EM, Bullaughey K, Matute DR, et al. (2012) Revisiting an old riddle: what determines genetic diversity levels within species? . *PLoS Biology* **10(9)** e1001388.
- Liersch S, Schmid-Hempel P (1998) Genetic variation within social insect colonies reduces parasite load. *Proceedings of the Royal Society B: Biological Sciences* **265**, 221-225.
- Lourenço A, Álvarez D, Wang IJ, Velo-Antón G (2017) Trapped within the city: integrating demography, time since isolation and population-specific traits to assess the genetic effects of urbanization. *Molecular Ecology* **26(6)**, 1498-1514.
- Lozier JD (2014) Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. *Molecular Ecology* **23**, 788-801.
- Lubin Y, Bilde T (2007) The evolution of sociality in spiders. *Advances in the Study of Behavior* **37**, 83-145.
- Lukas D, Clutton-Brock T (2012) Cooperative breeding and monogamy in mammalian societies. *Proceedings of the Royal Society B-Biological Sciences* **279**, 2151-2156.
- Massonnet B, Simon JC, Weisser WW (2002) Metapopulation structure of the specialized herbivore *Macrosiphoniella tanacetaria* (Homoptera, Aphididae). *Molecular Ecology* **11**, 2511-2521.
- Mattila TM, Bechsgaard JS, Hansen TT, Schierup MH, Bilde T (2012) Orthologous genes identified by transcriptome sequencing in the spider genus *Stegodyphus*. *BMC Genomics* **13**, 70.
- Nomura T, Takahashi J (2012) Effective population size in eusocial Hymenoptera with worker-produced males. *Heredity* **109**, 261-268.
- Nordborg M (2000) Linkage Disequilibrium, Gene Trees and Selfing: An Ancestral Recombination Graph With Partial Self-Fertilization. *Genetics* **154**, 923-929.
- Pannell JR, Charlesworth B (2000) Effects of metapopulation processes on measures of genetic diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **355**, 1851-1864.
- Pike N, Whitfield JA, Foster WA (2007) Ecological correlates of sociality in *Pemphigus* aphids, with a partial phylogeny of the genus. *Bmc Evolutionary Biology* **7**, 185.
- Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *Plos One* **7(2)**, e32253.

- Pusey A, Wolf M (1996) Inbreeding avoidance in animals. *Trends in Ecology & Evolution* **11**, 201-206.
- Pusey AE (1987) Sex-biased dispersal and inbreeding avoidance in birds and mammals. *Trends in Ecology & Evolution* **2**, 295-299.
- R Development Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics* **197**, 573-589.
- Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology* **17**, 230-237.
- Reeve HK, Westneat DF, Noon WA, Sherman PW, Aquadro CF (1990) DNA fingerprinting reveals high-levels of inbreeding in colonies of the eusocial naked mole-rat. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 2496-2500.
- Romiguier J, Gayral P, Ballenghien M, *et al.* (2014a) Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**, 261-263.
- Romiguier J, Lourenco J, Gayral P, *et al.* (2014b) Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *Journal of Evolutionary Biology* **27**, 593-603.
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.
- Sanggaard KW, Bechsgaard JS, Fang XD, *et al.* (2014) Spider genomes provide insight into composition and evolution of venom and silk. *Nature Communications* **5**, 3765.
- Schneider JM, Roos J, Lubin Y, Henschel JR (2001) Dispersal of *Stegodyphus dumicola* (Araneae, Eresidae): They do balloon after all! *Journal of Arachnology* **29**, 114-116.
- Schoen DJ, Brown AH (1991) Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences, United States of America* **88**, 4494-4497.
- Schou MF, Loeschcke V, Kristensen TN (2015) Inbreeding depression across a nutritional stress continuum. *Heredity* **115**, 56-62.
- Schrider DR, Shanku AG, Kern AD (2016) Effects of Linked Selective Sweeps on Demographic Inference and Model Selection. *Genetics* **204**, 1207-1223.
- Settepani V, Bechsgaard J, Bilde T (2014) Low genetic diversity and strong but shallow population differentiation suggests genetic homogenization by metapopulation dynamics in a social spider. *Journal of Evolutionary Biology* **27**, 2850-2855.
- Settepani V, Bechsgaard J, Bilde T (2016) Phylogenetic analysis suggests that sociality is associated with reduced effectiveness of selection. *Ecology and Evolution* **6**, 469-477.
- Spielman D, Brook BW, Briscoe DA, Frankham R (2004a) Does inbreeding and loss of genetic diversity decrease disease resistance? *Conservation Genetics* **5**, 439-448.
- Spielman D, Brook BW, Frankham R (2004b) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences, United States of America* **101**, 15261-15264.
- Stebbins GL (1957) Self fertilization and population variability in the higher plants. *American Naturalist* **91**, 337-354.
- Tabadkani SM, Nozari J, Lihoreau M (2012) Inbreeding and the evolution of sociality in arthropods. *Naturwissenschaften* **99**, 779-788.
- Taborsky M (2009) Reproductive skew in cooperative fish groups: virtue and limitations of alternative modelling approaches. . In: *Reproductive Skew in Vertebrates: Proximate and Ultimate Causes* (eds. Hager R, Jones CB), pp. 265-304. Cambridge University Press., Cambridge, United Kingdom.
- Takebayashi N, Morrell PL (2001) Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *American Journal of Botany* **88**, 1143-1150.

- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.
- Thorne BL (1997) Evolution of eusociality in termites. *Annual Review of Ecology and Systematics* **28**, 27-54.
- Ulbrich K, Henschel JR (1999) Intraspecific competition in a social spider. *Ecological Modelling* **115**, 243-251.
- Wade MJ, McCauley DE (1988) Extinction and recolonization - Their effects on the genetic differentiation of local populations. *Evolution* **42**, 995-1005.
- Walser B, Haag CR (2012) Strong intraspecific variation in genetic diversity and genetic differentiation in *Daphnia magna*: the effects of population turnover and population size. *Molecular Ecology* **21**, 851-861.
- Wang JL, Hill WG, Charlesworth D, Charlesworth B (1999) Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate. *Genetical Research* **74**, 165-178.
- West SA (2009) Sex allocation. In: *Monograph in Population Biology* (eds. Levin SA, Horn HS), Princeton University Press.
- Whitehorn PR, Tinsley MC, Brown MJF, Darvill B, Goulson D (2011) Genetic diversity, parasite prevalence and immunity in wild bumblebees. *Proceedings of the Royal Society B: Biological Sciences* **278**, 1195-1202.
- Whitlock MC, McCauley DE (1990) Some Population Genetic Consequences of Colony Formation and Extinction: Genetic Correlations within Founding Groups. *Evolution* **44**, 1717-1724.
- Wilson EO (1971) *The Insect Societies* Harvard University Press, Cambridge.
- World Spider Catalog (2017) *The World Spider Catalog, Version 18*. Natural History Museum Bern
- Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**, 0097-0159.
- Wright SI, Kalisz S, Slotte T (2013) Evolutionary consequences of self-fertilization in plants. *Proceedings of the Royal Society B-Biological Sciences* **280**, 20130133.
- Wright SI, Ness RW, Foxe JP, Barrett SCH (2008) Genomic consequences of outcrossing and selfing in plants. *International Journal of Plant Sciences* **169**, 105-118.

Figure legends

Figure 1. Geographic distribution of the populations sampled. Social species are underlined and represented by circles, subsocial sister species are represented by triangles and the subsocial outgroup is represented by a square. For a phylogeny of the genus see (Settepani *et al.* 2016). Full names of sampling localities and GPS points are provided in Table S1.

Figure 2. A) Observed individual heterozygosity (H_{obs}) and B) genetic diversity per population (π_S) and species (π_T). Error bars represent 95% confidence intervals. Phylogenetic relationships including posterior node probabilities and split time estimates represented on the left side of the graph are redrawn from Settepani *et al.* (2016). Social species are underlined. *S. dufuori*, *S. bicolor* and *S. tibialis* are not subject to investigation in this study, but are drawn to illustrate the multiple speciation events occurring from subsocial species. A) Individual heterozygosity was calculated as the mean frequency of polymorphic sites per individual per species. Individual estimates were obtained in pyRAD. B) Genetic diversity per species (π_T) is represented by white-striped bars, black-striped bars represent the *S. mimosarum* lineages (Malagasy and South African) and solid bars represent genetic diversity per population (π_S). Values of θ and Tajima's D per population and species are visible in Fig. S3.

Figure 3. Estimated population structure and F_{st} measures. Phylogenetic relationships including posterior node probabilities and split time estimates represented on the left side of the graph are redrawn from Settepani *et al.* (2016). Social species are underlined. *S. dufuori*, *S. bicolor* and *S. tibialis* are not subject to investigation in this study, but are drawn to illustrate the multiple speciation events occurring from subsocial species. A) Population structure (InStruct): The best representing number of genetic clusters (K) estimated for each species is reported under the species name and graphically represented by the barplots. Different colours in the barplot represent individuals' membership to different genetic clusters and inbreeding coefficients (F_{IS}) per cluster are shown on top of each cluster. Geographic population names are labelled below the barplot (see Fig. 1 for geographic localities). Results for the social *S. mimosarum* are reported for the species level and separately for the Malagasy and the South African clusters. A graphical representation of the best fastSTRUCTURE estimate of K per species is given in Fig. S4. B) Population differentiation (F_{st}): F_{st} was estimated by $1 - \pi_P/\pi_S$, with π_P being the average π of the populations under consideration, and π_S the total π of the species under consideration. F_{st} values were estimated for each species and separately for each genetic cluster of the social *S. mimosarum* (Madagascar and South Africa). Error bars represent 95% confidence intervals.

Figure 4. Demographic modelling from joint folded site frequency spectra (SFS). Two models were run for each species using fastsimcoal2 (Excoffier *et al.* 2013); one with and one without migration. The species to the left are the subsocial species and to the right their social sister-species. For each species the best fitting model based on Akaike Information Criterion (AIC) is shaded in grey. Below each model are population names and estimated effective population sizes (N_e) and 95 % confidence limits. The thickness of the branches is to scale with the N_e . To the left is a time scale for the time to the most recent common ancestor (TMRCA) in years. Note that the Y-axis is logarithmic. To see heatmaps representing joint folded site frequency spectra (SFS) see Fig. S2.